Convolution Neural Networks for Person Identification and Verification Using Steady State Visual Evoked Potential

Conference Paper · October 2018

DOI: 10.1109/SMC.2018.00188

CITATIONS
3

READS
108

6 authors, including:

Heba El-Fiqi
Zagazig University
14 PUBLICATIONS **27** CITATIONS
SEE PROFILE

Min Wang
UNSW Sydney
15 PUBLICATIONS **37** CITATIONS
SEE PROFILE

Nima Salimi
UNSW Australia
5 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Kathryn Kasmarik
Australian Defence Force Academy
114 PUBLICATIONS **717** CITATIONS
SEE PROFILE

2018 IEEE International Conference on Systems, Man, and Cybernetics

Convolution Neural Networks for Person Identification and Verification Using Steady State Visual Evoked Potential

Heba El-Fiqi, Min Wang, Nima Salimi, Kathryn Kasmarik, Michael Barlow, and Hussein Abbass
School of Engineering and IT
UNSW Canberra
Canberra, Australia
h.el-fiqi@adfa.edu.au,{Min.Wang,n.salimi}@student.adfa.edu.au,{k.kasmarik,m.barlow,h.abbass}@adfa.edu.au

***Abstract*—EEG signals could reveal unique information of an individual's brain activities. They have been regarded as one of the most promising biometric signals for person identification and verification. Steady-State Visual Evoked Potentials (SSVEPs), as EEG responses to visual stimulations at specific frequencies, could provide biometric information. However, current methods on SSVEP biometrics with hand-crafted power spectrum features and canonical correlation analysis (CCA) present only a limited range of individual distinctions and suffer relatively low accuracy.**

In this paper, we propose convolution neural networks (CNNs) with raw SSVEPs for person identification and verification without the need for any hand-crafted features. We conduct a comprehensive comparison between the performance of CNN with raw signals and a number of classical methods on two SSVEP datasets consisting of four and ten subjects, respectively. The proposed method achieved an averaged identification accuracy of $96.8\% \pm 0.01$, which outperformed the other methods by an average of 45.5% (p-value < 0.05). In addition, it achieved an averaged False Acceptance Rate (FAR) of $1.53\% \pm 0.01$ and True Acceptance Rate (TAR) of $97.09\% \pm 0.02$ for person verification. The averaged verification accuracy is $98.34\% \pm 0.01$, which outperformed the other methods by an average of 11.8% (p-value < 0.05). The proposed method based on deep learning offers opportunities to design a general-purpose EEG-based biometric system without the need for complex pre-processing and feature extraction techniques, making it feasible for real-time embedded systems.

I. Introduction

Recently, the use of Electroencephalography (EEG) signals as biometrics has received a great deal of attention. An EEG-based user recognition system has advantages over conventional biometrics data modalities such as face, fingerprint, and iris. EEG as data modality adheres to privacy compliance, security, and universality among other reasons [1].

Several EEG acquisition protocols have been proposed to develop automatic person recognition systems. Some studies are based on resting state EEG (eyes closed and eyes open) in which the user is not required to perform any mental task [2], [3]. Other studies investigated the effectiveness of mental tasks like motor imagery [4] and imagined speech production [5]. Another promising EEG protocol was where users are asked to perform mental tasks while they were presented with visual stimuli (e.g. black and white drawings). In this protocol, the recorded visual evoked potential (VEP) has been used for the purpose of individual identification [6].

Steady-state visual evoked potential (SSVEP) is a particular type of VEP that can be elicited when the subject focuses his/her attention on a repetitive visual stimulus [7]. The stimulus frequency and its harmonics can be detected from SSVEPs using signal processing and pattern recognition techniques. The repetitive visual stimulus can be a LED, a single graphic (e.g. square) on a screen, or a pattern reversal stimulus such as a checkerboard [8]. SSVEP has been applied in the development of brain-computer interfaces due to its high signal to noise ratio (SNR) and high bit rates [9]. Despite the successful applications of SSVEP in the development of BCIs, the usefulness of SSVEP in automatic user recognition has rarely been investigated.

SSVEP-based user recognition systems should leverage the advantages of well-established SSVEP protocols applied in other areas. Among a few pieces of work carried out in the area of SSVEP-based person recognition, Falzon et al. (2017) [10] successfully combined SSVEP recording protocols with the automatic user recognition requirements. In the corresponding study, SSVEP responses to six stimulus frequencies, ranging from 6.67 to 15 Hz, were recorded from eight subjects in three different sessions. The stimulus used in this study was a white square flickering on the black background of a 15.4-inch LCD display. The extracted feature vectors for classification consisted of data of all six frequencies and 32 channels (feature vectors with the length of 192). Principle component analysis (PCA) was applied to decrease the dimension of feature vectors to 15 principle components [11]. K-nearest neighbor (k-NN) classifier was used to identify the user based on computed feature vectors. It is noteworthy that in this study one session was used for the training and two other sessions were used for the testing purpose to evaluate the consistency of extracted features over an extended period of time (the second and third recording sessions were three and six weeks after the first session, respectively). The proposed method resulted in an averaged TAR of 91.7% and averaged FAR of 1.0%.

The successful application of classical machine learning algorithms such as k-NN requires a complex feature engi-

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00188

neering stage, particularly when dealing with a very dynamic data source like EEG. Deep learning which has revolutionized the image recognition research has also demonstrated some promising performance in the analysis of EEG signals. Deep learning has been used in EEG-based classification systems such as automatic emotion recognition [12] and motor imagery classification [13]. As deep learning algorithms can learn from raw time series, the complex feature extraction stage would be eliminated in the EEG signal processing. In this study, a user recognition system is proposed that combines SSVEP protocol with convolutional neural network (CNN).

II. Methodology

In this section, we present the proposed methods for SSVEP-based person identification and verification scenarios. The major difference between these two scenarios is whether a claimed identity is required with observations. An identification system outputs the identity of a person by checking the observations against all the persons in the database, and a verification system verifies whether the person is who he/she claims to be.

A. Convolutional Neural Networks

In our implementation, two different architectures have been designed for person identification and verification, respectively. For the identification problem, the architecture consists of three sets of (C-C-S) followed by a flatten fully connected layer (FC), then a Softmax multi-class classifier. C-C-S refers to using two convolutional layers (C) followed by a single Subsampling layer (S) and drop out. For the verification problem, two sets of (C-C-S) followed by a flatten fully connected layer, then a Sigmoid binary classifier. In summary, for the subject identification problem, the CNN network consists of (CCS-CCS-CCS-FC), and for the verification problem, the CNN network consists of(CCS-CCS-FC).

1) Input Layer: In our network, we have used the eight channels of the EEG signals as a 2D shaped input for the network. The input shape is 8 x 256 for each sample. As a pre-processing step, the training data has been standardized into zero mean and one standard deviation across all training examples.

2) Convolution layers: In Convolution layers, units are organized in planes which are known as feature maps. The way that the convolution layer works enable it to detect local conjunctions of features from the previous layer by applying different filters to extract the information. In our implementation, we have used small kernels of size 3 x 3 for the identification problem. For the verification problem, we have used kernels of size 3x3 for the first two convolution layers and 4x4 for the next two convolution layers. A different number of filters have been used across the layers for the two problems. The network architecture is shown in Fig. 1 and the detailed information about network configuration is described in Table I.

3) Non-linearity: We added non-linearity for the convolution layers by choosing non-linear activation functions for the neurons. We have used *Rectified Linear Unit* ReLU, which is half-wave non-linear function that map non-negative values to zeros. $f(x) = max(0, x)$

4) Subsampling: The sub-sampling step is used to reduce the dimensionality of each feature map spatially while maintaining the important characteristics of this feature map. That spatial pooling can be applied using aggregation functions like average, sum, max, etc. In our implementation, we used Max Pooling of 2 x 2 size window.

5) Dropout: Dropout have been used recently in deep neural network training for the purpose of reducing the network tendency to overfit during the training process [14]. Therefore, we used it after the first two sets of (C-C-S) layers, and again after the fully connected hidden layer.

6) Loss Function: The network has been trained using cross entropy loss. An *Adaptive Moment Estimation* optimization method, which is known as Adam optimizer [15] has been used to compute adaptive learning rate for each hyper-parameter. An initial learning rate value of 0.001 has been used.

For the verification problem, we have used binary cross entropy loss as in Equation 1.

$$BinaryLoss(o) = -(y_{o,c_1} log(p_{o,c_1}) + (1 - y_{o,c_1}) log(1 - p_{o,c_1})) \quad (1)$$

For the identification problem, we have changed to categorical cross entropy loss, shown in Equation 2 as it is a multi-class classification problem.

$$CategoricalLoss(o) = -\sum_{c=1}^{K} y_{o,c} log(p_{o,c}) \quad (2)$$

where K indicates the number of classes, and the log here refers to the natural log. y is one hot encoding for the correct class. $y_{o,c}$ equals 1 if the class label c is the actual class of the observation o, and $p_{o,c}$ represents the probability that observation o is predicted as class c. Equation 2 refers to the loss for the classification associated with single observation. For the entire dataset, we use Equation 3, where N is the number of all the observations included in the training set.

$$Loss = -\sum_{o=1}^{N} \sum_{c=1}^{K} y_{o,c} log(p_{o,c}) \quad (3)$$

For classification, the identification problem is a multi-class classification problem, therefore, we have used $softmax$ classifier, as in Equation 4, in the last layer of our architecture. w refers to the weight vector and x refers to the input vector of observation o

$$P(y = c|x) = \frac{e^{x^T w_c}}{\sum_{k=1}^{K} e^{x^T w_k}} \quad (4)$$

For the verification problem, it is a binary classification problem. Therefore,we use $Sigmoid$ classifier as in Equation 5.

$$P(y = c|x) = \frac{1}{1 + e^{-x^T w_c}} \quad (5)$$

TABLE I
CNN NETWORK SETTINGS FOR CONVOLUTION LAYERS, POOLING LAYERS, AND ACTIVATION LAYERS

CNN Configuration	Identification Problem	Verification Problem
Layers Sequence	C1-C2-S1-C3-C4-S2-C5-C6-S3-FC	C1-C2-S1-C3-C4-S2-FC
Kernel size used in Conv layers	3x3	3x3 and 4x4
Activation functions in Conv Layers	Relu	
Conv layers C1, C2	No. of Kernels=32	
Subsampling layers S1, S2	Pooling kernel size=2x2	
Conv layers C3, C4	No. of Kernels=64	
Conv layers C5, C6	No. of Kernels= 128	—
Subsampling layer S3	Pooling kernel size=2x2	—
Fully connected layer FC	256	128
Classifier	Softmax	Sigmoid

B. Signal Enhancement based on CCA

Canonical Correlation Analysis (CCA) has been a popular method for frequency recognition in SSVEP-based BCI systems [16], [17]. In this paper, we apply CCA as a spatial filter for SSVEP signal enhancement, proposed in [18], to increase the visibility of the underlying characteristics. The purpose is to investigate whether the CCA-based spatial filter is valid for SSVEP-based person recognition.

CCA is a multivariate statistical method that explores the underlying correlation between two signals. It seeks a pair of linear transforms $\mathbf{w}_x$ and $\mathbf{w}_y$ to maximise the correlation between the linear combinations $\hat{\mathbf{x}} = \mathbf{w}_x'\mathbf{X}$ and $\hat{\mathbf{y}} = \mathbf{w}_y'\mathbf{Y}$ according to Equation 6.

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \rho = \frac{\mathbf{w}_x' C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{xx} \mathbf{w}_x \mathbf{w}_y' C_{yy} \mathbf{w}_y}} \tag{6}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are two given sets of signals; ρ is the maximal canonical correlation; C_{xy}, C_{xx} and C_{yy} are inter and intra-sets covariance matrix, respectively.

In the case of signal enhancement, CCA is applied to SSVEP signals as a spatial filter to improve the signal-to-noise ratio. Assume $\mathbf{X} \in \mathbf{R}^{C\times N}$ is a training dataset which consists of EEG signals from C channels with N time points in each channel. We apply CCA to find the maximal correlation between the EEG signal set $\mathbf{X}$ and a set of reference signals $\mathbf{Y}$, where $\mathbf{Y} \in \mathbf{R}^{2H\times N}$ is the pre-constructed reference signal set which is defined by a series of sine-cosine waves at the stimulation frequency $f_m \quad (m = 1, 2, \ldots, M)$, according to Equation 7.

$$\mathbf{Y}_m = \begin{bmatrix} \sin(2\pi f_m t) \\ \cos(2\pi f_m t) \\ \vdots \\ \sin(2\pi H f_m t) \\ \cos(2\pi H f_m t) \end{bmatrix} \quad t = \frac{1}{Fs}, \frac{2}{Fs}, \ldots, \frac{N}{Fs} \tag{7}$$

where Fs is the signal sampling rate and H is the harmonics number. Then we calculate the correlation between $\mathbf{X}$ and $\mathbf{Y}_m$, and save the projection weight vector $\mathbf{w}_x$ which yields the maximal ρ with all EEG samples. The optimized weights are used for constructing the output of the spatial filter, which is a single time-series.

III. EXPERIMENTAL DESIGN

In this section, we present the design of the experimental analysis that includes the datasets and preprocessing techniques, the evaluation procedures to compare deep neural networks with shallow ones and conventional classification criteria, and the experiments that we conducted in this analysis.

A. Datasets and Preprocessing

The SSVEP signals for testing our proposed methods were gathered from two public datasets [18], [19]. The two datasets consist of EEG recordings from 5 and 12 healthy subjects, respectively, and their age range is between 20 and 30. Similar SSVEP signal acquisition protocol and experimental conditions are adopted for both datasets. For SSVEP stimulation, LED arrays were chosen to provide flash stimuli instead of computer screens to avoid the effects of self-refreshing rate. A micro-controller was set up to control the three groups of LED arrays blinking at three frequencies (i.e. 13Hz, 17Hz and 21Hz), accordingly. Subjects were seated comfortably on a reclining chair and allowed to move and blink freely. EEG signals were recorded from eight channels in parietal-occipital areas at 256Hz while the subjects focusing on the flashing LED stimuli, and recorded EEG were referenced to the mean signal from earlobes. The electrode placements of the eight channels are illustrated in Fig. 2. The whole recording process consists of multiple sessions for each subject on different days, with the same operation and in the same conditions. We deleted the bad sessions, which are reported in the datasets to be invalid due to synchronization problem and hardware failure. We excluded one subject from dataset I according to the description of the dataset, and also two subjects from dataset II were excluded due to the bad epochs.

A Hamming window sinc FIR filter was applied to restrict available frequency range to [10 40] Hz. For artifacts rejection, we conducted ICA and reconstructed signals after removing contaminated components identified to be responsible for ocular and muscular artifacts by MARA algorithm [20]. A moving window of 1 second with 50% overlap was adopted for generating classification samples for evaluating the methods. These samples are labeled with subject ID (i.e. 1 to 4 for dataset I and 1 to 10 for dataset II). Detailed information about the sample size of each subject in the two datasets are summarised in Table II.

TABLE II
SAMPLE SIZE OF EACH DATASET

	Number of subject	Number of classification samples per subject
Dataset I	4	840
Dataset II	10	312

B. Evaluation Procedures

For performance evaluation, 15 runs of training and testing were conducted for each method. In each run, we randomly

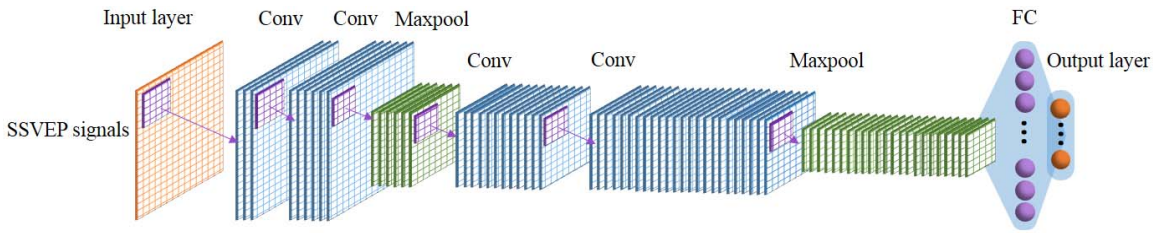

Fig. 1. Deep architecture design for the proposed CNN.

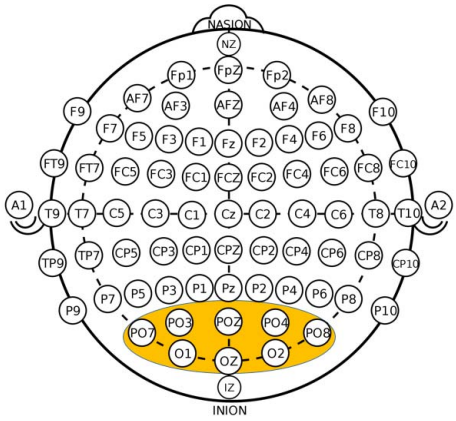

Fig. 2. Electrode placements of the 10-20 system. The eight electrodes used in the experiment are highlighted in yellow.

picked 80% of the samples for training and used the remaining 20% for testing. A fixed set of 15 different random seeds was adopted for generating these 15 training-testing sets. This evaluation procedure extended the standard 5-fold cross-validation by providing enough runs to enable a significant test over the methods. For deep learning method, we randomly picked 10% samples from the training set as a validation set and applied early stopping for training procedure based on the error evaluated on the validation set. This error is used as a proxy for the generalization error to avoid over-fitting and prevent poor generalization performance.

As identification is a one-against-all classification task, we calculated correct identification accuracy for performance evaluation. For verification, FAR and TAR were utilized in addition to accuracy as they are common measures for evaluating verification systems. FAR is the measure of the likelihood that the system will incorrectly accept an unauthorized user; while TAR reflects the likelihood that a genuine user will be accepted. The two metrics are calculated according to the following equations.

$$FAR = \frac{FP}{(FP + TN)} \tag{8}$$

$$TAR = 1 - \frac{FN}{(TP + FN)} \tag{9}$$

These two measures are basic metrics for assessing or comparing the performance of biometric verification systems in regards to security level and user convenience. Specifically, the aim is to increase TAR, therefore, to enhance user convenience, while decreasing FAR for higher security level.

C. Experiments

We investigated the state-of-the-art SSVEP classification methods for person recognition and compared our proposed methods with each of them. The first comparison method is the classical SSVEP classification criterion, which is applying conventional classifiers on spectral features extracted from the SSVEP signals. The second comparison method is applying CCA spatial filter to the raw SSVEP time-series before extracting power spectral features for classification. Three classical classifiers were selected for comparison, namely random forest (RF), support vector machine (SVM) and a shallow feed-forward neural network (NN) with one hidden layer.

Our proposed deep learning modal is working directly on the raw time-series without any feature extraction processing, which provides feasibility of real-time applications. In order to evaluate the learning ability of the neural networks, we also compared our deep learning model with the shallow neural network using signal time-series input without extracting any feature. Fig. 3 depicts the working procedure of the proposed method and the comparison methods. For clarity, a code is assigned to each method as follows:

- Conventional classifiers on power spectrum features extracted from raw signals: PSF-RF, PSF-SVM and PSF-NN.
- Conventional classifiers on power spectrum features after CCA processing: CCA-PSF-RF, CCA-PSF-SVM and CCA-PSF-NN.
- Shallow neural network on raw signals: Raw-NN.
- Convolutional neural network on raw signals and CCA-enhanced signals: Raw-CNN, CCA-CNN.

IV. Results and Discussion

The person identification performance of the proposed method and comparison methods are presented in Table III and Table V for Dataset I and II, respectively. For each method, we present the averaged identification accuracy and standard deviation across 15 runs. It can be seen from the table that our proposed method, Raw-CNN, achieved the best performance among all the methods, providing an average of 97.60% and 95.90% for Dataset I and II, respectively. The standard deviation across runs is around 0.007 and 0.01 for Dataset I

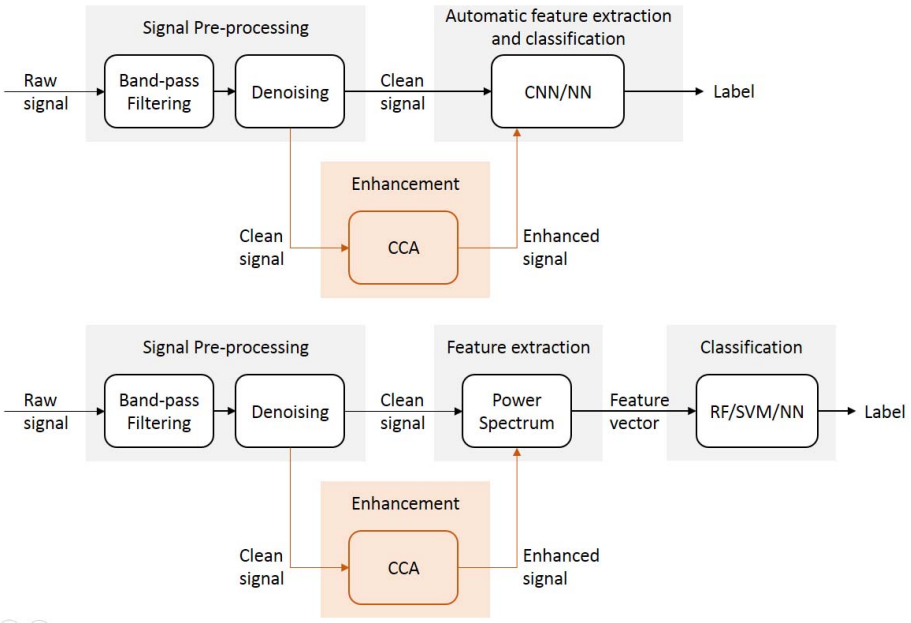

Fig. 3. Workflow of the proposed method and all the comparison methods

TABLE III
RESULTS FOR PERSON IDENTIFICATION PROBLEM FOR DATASET I

Identification accuracy (%) over 15 runs									
	With CCA filtering				Raw SSVEP time-series				
Results	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN	Raw-CNN
Average	57.77%	72.87%	63.54%	55.26%	73.58%	91.28%	90.01%	51.06%	**97.60**%
STD	1.9E-02	1.6E-02	1.8E-02	1.8E-02	1.5E-02	1.0E-02	9.1E-03	1.3E-02	7.0E-03

TABLE IV
T-TEST RESULTS: PERSON IDENTIFICATION PROBLEM FOR DATASET I (RAW-CNN VS ALTERNATIVE METHODS)

	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN
p-value	1.6E-19	2.9E-18	4.6E-19	7.6E-21	1.2E-17	1.5E-11	1.6E-12	2.8E-22
decision	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0

TABLE V
RESULTS FOR PERSON IDENTIFICATION PROBLEM FOR DATASET II

Identification accuracy (%) over 15 runs									
	With CCA filtering				Raw SSVEP time-series				
Results	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN	Raw-CNN
Average	58.47%	69.75%	59.37%	57.67%	69.17%	93.34%	89.97%	53.78%	**95.90**%
STD	1.5E-02	1.7E-02	1.4E-02	1.9E-02	1.6E-02	8.9E-03	9.8E-03	1.9E-02	1.1E-02

and II, indicating the performance is consistent. Additionally, by comparing results between the two datasets, we can see that Raw-CNN provides more robust performance over different subject datasets, while the other methods are more sensitive to different subjects. For example, PSF-SVM achieved an averaged identification accuracy of 93.34% on Dataset II. However, the accuracy dropped to 91.28% on Dataset I which is not acceptable. In order to investigate whether the proposed method, Raw-CNN, significantly improves the identification accuracy, a one-tailed paired t-test (significant level 0.05) was conducted on the results of Raw-CNN against each alternative method. The null hypothesis is that the population mean of Raw-CNN results is equal to the population mean of results obtained from each alternative method. The p-values and the corresponding decisions are presented in Table IV and Table VI. All the p-values are less than 0.05, indicating that our method based on CNN on raw EEG data significantly improves the identification performance in comparison with all the other methods. Combination of power spectral-based features with SVM and NN showed the best results among conventional

TABLE VI
T-TEST RESULTS: PERSON IDENTIFICATION PROBLEM FOR DATASET II (RAW-CNN VS ALTERNATIVE METHODS)

	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN
p-value	5.8E-20	1.8E-17	3.8E-21	2.4E-19	4.2E-17	2.3E-06	5.8E-10	5.3E-19
decision	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0

TABLE VII
RESULTS FOR PERSON VERIFICATION PROBLEM FOR DATASET I

Verification accuracy (%), FAR (%) and TAR (%)						
Metric	Method	subject 1	subject 2	subject 3	subject 4	Average
Accuracy	CCA-PSF-RF	80.37%±0.01	83.23%±0.01	77.19%±0.01	81.13%±0.02	80.48%±0.03
	CCA-PSF-SVM	86.41%±0.01	87.66%±0.01	82.12%±0.01	87.35%±0.02	85.89%±0.03
	CCA-PSF-NN	86.30%±0.01	86.28%±0.02	82.34%±0.01	87.54%±0.01	85.62%±0.02
	CCA-CNN	73.38%±0.07	39.78%±0.21	72.84%±0.06	54.44%±0.25	60.11%±0.16
	PSF-RF	83.08%±0.01	87.58%±0.01	80.33%±0.02	89.11%±0.01	85.03%±0.04
	PSF-SVM	94.52%±0.01	96.00%±0.01	93.46%±0.01	96.32%±0.01	95.08%±0.01
	PSF-NN	95.61%±0.01	97.14%±0.01	94.41%±0.01	97.36%±0.00	96.13%±0.01
	Raw-NN	81.67%±0.36	81.54%±0.30	77.68%±0.02	82.71%±0.01	80.90%±0.02
	Raw-CNN	**97.86%±0.01**	**98.36%±0.01**	**97.26%±0.02**	**98.93%±0.01**	**98.10%±0.01**
FAR	CCA-PSF-RF	4.74% ±0.01	5.13% ±0.01	4.15% ±0.01	3.58% ±0.01	4.40% ±0.01
	CCA-PSF-SVM	7.49% ±0.01	5.98% ±0.01	10.06% ±0.01	5.16% ±0.01	7.17% ±0.02
	CCA-PSF-NN	8.08% ±0.02	7.97% ±0.03	10.34% ±0.02	6.83% ±0.02	8.31% ±0.02
	CCA-CNN	5.50% ±0.16	75.29% ±0.39	7.09% ±0.15	47.05% ±0.45	33.73% ±0.34
	PSF-RF	3.40% ±0.01	3.54% ±0.01	3.51% ±0.01	2.32% ±0.01	3.19% ±0.01
	PSF-SVM	2.62% ±0.01	2.22% ±0.00	3.66% ±0.01	1.12% ±0.00	2.41% ±0.01
	PSF-NN	2.45% ±0.01	1.77% ±0.01	3.46% ±0.01	1.33% ±0.00	2.25% ±0.01
	Raw-NN	52.49% ±0.10	36.99% ±0.09	44.81% ±0.04	45.05% ±0.03	44.84% ±0.06
	Raw-CNN	**1.87% ±0.01**	**1.74% ±0.01**	**3.12% ±0.03**	**1.08% ±0.01**	**1.95% ± 0.01**
TAR	CCA-PSF-RF	36.18% ±0.04	49.65% ±0.04	20.19% ±0.04	34.40% ±0.04	35.10% ±0.12
	CCA-PSF-SVM	68.33% ±0.04	69.26% ±0.03	58.21% ±0.03	64.46% ±0.03	65.06% ±0.05
	CCA-PSF-NN	69.60% ±0.05	69.71% ±0.05	60.00% ±0.06	70.31% ±0.04	67.40% ±0.05
	CCA-CNN	10.01% ±0.21	83.02% ±0.32	11.64% ±0.24	59.41% ±0.37	41.02% ±0.36
	PSF-RF	42.83% ±0.04	61.96% ±0.03	30.93% ±0.05	62.92% ±0.03	49.66% ±0.16
	PSF-SVM	86.10% ±0.04	90.86% ±0.02	84.73% ±0.03	88.51% ±0.03	87.55% ±0.03
	PSF-NN	89.86% ±0.03	94.02% ±0.02	87.97% ±0.03	93.38% ±0.02	91.31% ±0.03
	Raw-NN	8.52% ±0.16	3.03% ±0.13	11.54% ±0.02	4.97% ±0.01	7.01% ±0.04
	Raw-CNN	**97.05% ±0.01**	**98.67% ±0.01**	**98.41% ±0.02**	**98.95% ±0.01**	**98.27% ± 0.01**

TABLE VIII
T-TEST RESULTS: PERSON VERIFICATION PROBLEM FOR DATASET I (RAW-CNN VS ALTERNATIVE METHODS)

	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN
p-value	2.10E-04	5.80E-04	3.10E-04	9.90E-03	2.20E-03	1.40E-03	6.10E-03	1.10E-04
decision	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0	reject H0

methods, and their performance was relatively comparable to that of Raw-CNN method. However, our proposed method based on CNN works directly on raw EEG signals, which reduces the feature extraction stage, therefore, can potentially improve the system operation speed.

For person verification, the accuracy, FAR and TAR performance of each method is presented in Table VII and Table IX for Dataset I and Dataset II, respectively. For each method, we present the averaged values and standard deviations of 15 runs for each subject, and the last column shows the averaged performance across all the subjects. It can be seen that for Dataset I, our proposed method Raw-CNN achieved the best performance in regards to verification accuracy, FAR and TAR. For Dataset II, Raw-CNN achieved an overall best accuracy and TAR. Although it didn't achieve the best FAR for this dataset, it is worth noting that all three metrics (FAR, TAR and accuracy) should be considered while comparing the verification performance of the methods. For Example, CCA-CNN method achieved best FAR of zeros for both Subjects 1 and 3. The CNN classified all the samples into the same class; thus, it didn't actually learn anything, TAR also shows zeros for these two subjects. Firstly, Raw-CNN achieved the highest verification accuracy for most of the individuals in the two datasets. Specifically, an averaged accuracy of 98.10% and 98.58% across subjects was achieved on Dataset I and Dataset II, respectively, which significantly improves the verification accuracy. Secondly, Raw-CNN provided the lowest FAR for each individual in Dataset I with an average of 1.95%. The FAR of Raw-CNN in Dataset II is slightly higher than some of the comparison methods, however, within acceptance range with an average of 1.11% across ten subjects. This indicates the proposed method provides the high-security level, that the likelihood of the system getting intruded by an unauthorized user is very low. Also, it is worth noticing that in some cases,

TABLE IX
RESULTS FOR PERSON VERIFICATION PROBLEM FOR DATASET II

Metric	Method	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6	subject 7	subject 8	subject 9	subject 10	Average
Accuracy	CCA-PSF-RF	98.24%±0.01	93.63%±0.01	94.14%±0.01	93.96%±0.01	93.43%±0.01	92.90%±0.01	93.25%±0.01	93.82%±0.01	93.17%±0.01	94.17%±0.01	94.07%±0.02
	CCA-PSF-SVM	99.15%±0.00	93.50%±0.01	95.79%±0.01	93.10%±0.01	91.13%±0.01	92.89%±0.01	93.10%±0.01	92.76%±0.01	93.20%±0.01	96.30%±0.01	94.09%±0.02
	CCA-PSF-NN	99.00%±0.00	93.50%±0.01	96.33%±0.01	94.45%±0.01	93.52%±0.01	93.56%±0.01	93.00%±0.01	94.44%±0.01	94.03%±0.01	96.68%±0.01	94.85%±0.02
	CCA-CNN	89.61%±0.01	90.58%±0.01	89.96%±0.01	91.21%±0.02	90.29%±0.01	89.87%±0.01	89.60%±0.01	90.83%±0.01	89.84%±0.01	90.41%±0.05	90.22%±0.01
	PSF-RF	98.61%±0.00	94.02%±0.01	94.34%±0.01	94.36%±0.01	93.57%±0.01	93.53%±0.01	93.25%±0.01	94.68%±0.01	94.00%±0.01	95.41%±0.01	94.58%±0.02
	PSF-SVM	99.43%±0.00	**98.18%± 0.01**	**99.55% ± 0.00**	99.17%±0.00	**96.95% ± 0.01**	97.73%±0.01	97.21%±0.00	**99.58% ± 0.00**	97.97%±0.01	99.36%±0.00	98.51%±0.01
	PSF-NN	99.51%±0.00	97.98%±0.01	99.24%±0.00	99.16%±0.00	96.78%±0.01	97.64%±0.01	97.41%±0.01	99.72%±0.00	97.96%±0.01	99.11%±0.01	98.45%±0.01
	Raw-NN	93.58%±0.01	93.29%±0.01	93.54%±0.01	93.60%±0.01	92.98%±0.01	92.47%±0.01	92.83%±0.01	93.41%±0.01	93.16%±0.01	94.33%±0.01	93.32%±0.01
	Raw-CNN	**99.62% ±0.01**	97.84%± 0.01	98.29%±0.01	**99.73%± 0.00**	96.62%±0.01	**99.50% ± 0.00**	**97.64% ± 0.02**	99.29%±0.01	**98.06% ± 0.02**	**99.25% ± 0.01**	**98.58% ± 0.01**
FAR	CCA-PSF-RF	0.27%±0.00	0.18%±0.00	0.26%±0.00	**0.11%±0.00**	**0.13%±0.00**	0.17%±0.00	**0.13%±0.00**	0.20%±0.00	0.31%±0.00	0.14%±0.00	0.19%±0.00
	CCA-PSF-SVM	0.23%±0.00	0.70%±0.00	1.92%±0.01	0.94%±0.00	0.04%±0.00	0.14%±0.00	0.45%±0.00	1.87%±0.00	1.69%±0.00	1.17%±0.00	0.91%±0.01
	CCA-PSF-NN	0.37%±0.00	2.18%±0.01	1.92%±0.01	1.71%±0.01	1.47%±0.01	1.35%±0.01	2.07%±0.01	2.47%±0.01	2.19%±0.01	1.45%±0.01	1.72%±0.01
	CCA-CNN	0.00%±0.00*	0.54%±0.01	0.00%±0.00*	0.90%±0.01	1.46%±0.02	0.50%±0.01	0.25%±0.01	0.74%±0.01	2.51%±0.01	0.08%±0.07	0.70%±0.01
	PSF-RF	0.32%±0.00	**0.12%±0.00**	**0.05%±0.00**	0.12%±0.00	0.14%±0.00	**0.07%±0.00**	0.24%±0.00	0.06%±0.00	**0.25%±0.00**	**0.06%±0.00**	**0.14% ± 0.00**
	PSF-SVM	0.38%±0.00	0.47%±0.00	0.16%±0.00	0.26%±0.00	0.70%±0.00	0.36%±0.00	0.29%±0.00	**0.02%±0.00**	0.73%±0.00	0.14%±0.00	0.35%±0.00
	PSF-NN	0.31%±0.00	0.78%±0.00	0.30%±0.00	0.36%±0.00	0.96%±0.00	0.68%±0.00	0.48%±0.00	0.13%±0.00	0.86%±0.00	0.54%±0.01	0.54%±0.00
	Raw-NN	**0.12%±0.00**	0.82%±0.00	0.74%±0.00	1.06%±0.00	1.21%±0.01	1.27%±0.01	0.93%±0.00	0.92%±0.00	1.07%±0.01	0.76%±0.00	0.89%±0.00
	Raw-CNN	0.27%±0.00	2.04%±0.01	1.42%±0.01	0.19%±0.00	2.43%±0.01	0.37%±0.00	1.35%±0.01	0.71%±0.01	1.71%±0.02	0.61%±0.01	1.11%±0.01
TAR	CCA-PSF-RF	85.47%±0.05	38.27%±0.05	44.04%±0.05	41.03%±0.05	37.21%±0.05	32.46%±0.06	36.88%±0.04	38.82%±0.06	37.44%±0.04	44.86%±0.04	43.65%±0.15
	CCA-PSF-SVM	93.69%±0.03	41.64%±0.05	75.58%±0.04	39.80%±0.06	14.01%±0.04	32.17%±0.06	38.25%±0.04	43.20%±0.05	49.48%±0.06	74.56%±0.04	50.24%±0.24
	CCA-PSF-NN	93.57%±0.03	55.32%±0.07	80.93%±0.04	60.23%±0.07	49.56%±0.07	49.48%±0.05	51.10%±0.06	66.10%±0.07	61.75%±0.08	80.61%±0.03	64.86%±0.15
	CCA-CNN	0.00%±0.00*	10.36%±0.18	0.00%±0.00*	21.63%±0.28	18.74%±0.21	6.41%±0.17	2.94%±0.11	13.18%±0.11	23.97%±0.19	8.36%±0.27	10.56%±0.09
	PSF-RF	89.32%±0.03	41.59%±0.05	44.07%±0.05	45.20%±0.05	38.66%±0.05	37.80%±0.05	37.75%±0.04	46.45%±0.07	44.85%±0.05	56.23%±0.07	48.19%±0.15
	PSF-SVM	97.81%±0.01	86.14%±0.04	**97.00% ± 0.03**	94.13%±0.03	76.58%±0.07	81.18%±0.04	75.80%±0.03	96.01%±0.02	86.87%±0.05	95.10%±0.03	88.66%±0.09
	PSF-NN	98.02%±0.01	86.94%±0.04	95.20%±0.03	94.80%±0.02	77.11%±0.04	83.04%±0.05	79.26%±0.05	98.41%±0.02	87.84%±0.04	96.16%±0.03	89.68%±0.08
	Raw-NN	39.44%±0.06	40.66%±0.05	42.31%±0.05	45.92%±0.05	42.33%±0.05	37.96%±0.06	39.73%±0.03	41.29%±0.05	44.09%±0.07	51.75%±0.04	42.55%±0.04
	Raw-CNN	**98.51% ± 0.04**	**96.91% ± 0.03**	95.64%±0.03	**99.04% ± 0.02**	**88.21%± 0.07**	**98.36%± 0.02**	**89.10%± 0.09**	**99.26%± 0.01**	**95.97%± 0.04**	**98.09% ± 0.03**	**95.91% ± 0.04**

TABLE X
T-TEST RESULTS: PERSON VERIFICATION PROBLEM FOR DATASET II (RAW-CNN VS ALTERNATIVE METHODS)

	CCA-PSF-RF	CCA-PSF-SVM	CCA-PSF-NN	CCA-CNN	PSF-RF	PSF-SVM	PSF-NN	Raw-NN
p-value	2.10E-06	3.10E-05	2.60E-05	6.10E-10	2.90E-06	3.90E-01	2.90E-01	2.00E-08
decision	reject H0	reject H0	reject H0	reject H0	reject H0	accept H0	accept H0	reject H0

e.g. for CCA-CNN in subject 1 and subject 3, the FAR and TAR is zero. The zeros do not reflect the true performance, they are due to the fact that the method classified all the samples as either genuine user or intruder. Finally, in regards to TAR, the Raw-CNN showed the highest TAR for each individual, achieving an average of 98.27% and 95.91% across subjects for Dataset I and II, respectively. The highest TAR value indicated that the Raw-CNN could potentially provide more user-friendly verification system, where the system has very low likelihood to reject the genuine users. Besides, the standard deviations of Raw-CNN across subjects in regard of all the three assessment metrics (verification accuracy, FAR and TAR) are very low, in some cases, they are even several orders of magnitude smaller than the other methods. This piece of result indicated that the Raw-CNN provides more robust verification performance across different users.

We conducted one-tailed paired t-test (significant level 0.05) on the verification accuracy results to validate the performance of the proposed method, Raw-CNN, against all the alternative methods. The null hypothesis assumes that there is no difference between the population mean results obtained from Raw-CNN and population mean results computed by each comparison method. The p-values and the corresponding decisions are presented in Table VIII and Table X. For Dataset I, all the p-values are less than 0.05. Therefore, the null hypothesis has been rejected. It indicated that our method based on CNN on raw EEG data significantly improves the performance. For Dataset II, we rejected all the null hypothesis, except for PSF-SVM and PSF-NN, with a significant level of 0.05. However, when comparing results from both Dataset I and Dataset II together, a significant improvement was observed for Raw-CNN against PSF-SVM and PSF-NN. The results indicated that Raw-CNN provides robust performance over different user pool, while PSF-SVM and PSF-NN are more sensitive to different users.

As mentioned earlier, the good performance of conventional learning algorithms is dependent on the feature engineering stage. Using only raw data as input to the NN (Raw-NN) did not result in better performance in compared to deep learning approach as shown by comparing the results of (Raw-NN) with (Raw-CNN) method for the two datasets. On the other hand, the better results for NN was achieved when FFT transformed data (spectral information) was used as input to the classifier (compare the results of (Raw-NN) with (PSF-NN) method in both datasets).

The obtained results in this study indicate the poor performance of CCA algorithm as a spatial filter technique, at least for the purpose of person recognition. Using CCA rather than raw data led to a decrease in the classification performance in both conventional and deep learning methods. One explanation for this results could be that CCA has filtered out discriminatory representations. Thus, use of any type of spatial filter (e.g. ICA, PCA, CCA) in the EEG-based user recognition development should be thoroughly investigated, particularly when deep learning approach is adopted.

Despite the promising performance of CNN in user recog-

nition application, the proposed model (Raw-CNN) needs to be further evaluated in more realistic experiments. Datasets used in this study were not recorded for the purpose of user recognition. Thus, it is necessary to test the proposed model with the data recorded for this purpose. For instance, data of different sessions, recorded at different weeks, should be used to train and test the model to assess the consistency of the proposed algorithm over time. Moreover, SVM and NN showed competitive performance, particularly for the bi-class (verification) problem. Combination of these two conventional classifiers with a better feature extractor (e.g. Short-time Fourier Transform (STFT), Wavelet Transform (WT)) might outperform the performance of our proposed model. Therefore, it is necessary to compare the (Raw-CNN) method with other combinations of feature extractors and classifiers in the future studies.

V. Conclusion

In this study, we proposed a deep learning model based on the convolutional neural network for analyzing SSVEP signals for person recognition, including two scenarios which are person identification and verification. In order to demonstrate the advantage of using deep learning, we also investigated two other criteria, including (1) adopting conventional classifiers on extracted power spectrum features; and (2) conducting CCA-based spatial filter on SSVEP signals before extracting features. We evaluated the methods by accuracy, FAR and TAR (for verification only) for performance assessment. Results showed that the proposed deep learning model based on convolutional neural networks significantly outperformed the conventional methods in terms of recognition performance and robustness, which indicated that it could automatically learn discriminative representations from raw SSVEP signals for person recognition without any feature extraction processing.

In the future extension of this work, we will evaluate the proposed method on larger datasets and conduct a permanence analysis of SSVEP-based person recognition system. Additionally, we will investigate the capability of the proposed method to deal with transfer learning problem across subjects.

Acknowledgment

The study was conducted after approval from UNSW Human Research Ethics Compliance Committee Protocol HC17434. This work was funded by the Australian Research Council Discovery Grant number DP160102037.

References

[1] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE transactions on information forensics and security*, vol. 9, no. 5, pp. 782–800, 2014.

[2] M. Poulos, M. Rangoussi, and N. Alexandris, "Neural network based person identification using EEG features," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 2. IEEE, 1999, pp. 1117–1120.

[3] D. La Rocca, P. Campisi, and G. Scarano, "On the repeatability of EEG features in a biometric recognition framework using a resting state protocol." in *BIOSIGNALS*, 2013, pp. 419–428.

[4] S. Marcel and J. d. R. Millán, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, 2007.

[5] K. Brigham and B. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–8.

[6] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[7] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottereau, and B. Rossion, "The steady-state visual evoked potential in vision research: a review," *Journal of vision*, vol. 15, no. 6, pp. 4–4, 2015.

[8] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," *Computational intelligence and neuroscience*, vol. 2010, p. 1, 2010.

[9] S. Parini, L. Maggi, A. C. Turconi, and G. Andreoni, "A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication," *Computational Intelligence and Neuroscience*, vol. 2009, 2009.

[10] O. Falzon, R. Zerafa, T. Camilleri, and K. P. Camilleri, "EEG-based biometry using steady state visual evoked potentials," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017, pp. 4159–4162.

[11] S. K. Goh, H. A. Abbass, K. C. Tan, and A. Al Mamun, "Artifact removal from eeg using a multi-objective independent component analysis model," in *International Conference on Neural Information Processing*. Springer, 2014, pp. 570–577.

[12] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[13] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *International Conference on Intelligent Computing*. Springer, 2014, pp. 203–210.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014. [Online]. Available: http://dl.acm.org/citation.cfm?id=2627435.2670313

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.

[16] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *International journal of neural systems*, vol. 24, no. 04, p. 1450013, 2014.

[17] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[18] E. Kalunga, K. Djouani, Y. Hamam, S. Chevallier, and E. Monacelli, "SSVEP enhancement based on canonical correlation analysis to improve BCI performances," in *AFRICON, 2013*. IEEE, 2013, pp. 1–5.

[19] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Monacelli, and Y. Hamam, "Online SSVEP-based BCI using riemannian geometry," *Neurocomputing*, vol. 191, pp. 55–68, 2016.

[20] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann, "Robust artifactual independent component classification for BCI practitioners," *Journal of neural engineering*, vol. 11, no. 3, p. 035013, 2014.